LA-UR-04-6564

Approved for public release; distribution is unlimited.

Title: INFORMATION INTEGRATION FOR COMPLEX SYSTEMS

Author(s): ALYSON G. WILSON, LANL
LAURA A. MCNAMARA, SNL
GREGORY D. WILSON, LANL

Submitted to:

Information Integration for Complex Systems

Alyson G. Wilson, Laura A. McNamara, and Gregory D. Wilson*

October 18, 2004

* **Alyson Wilson** is a Technical Staff Member and the Technical Lead for DoD Programs in the Statistical Sciences Group at Los Alamos National Laboratory, P.O. Box 1663, MS F600, Los Alamos, NM 87545 (email: *agw@lanl.gov*). Laura McNamara is a Principal Member of the Technical Staff in the Cooperative Monitoring Center at Sandia National Laboratories (email: *lamcnam@sandia.gov*). Gregory Wilson is a Technical Staff Member in the Statistical Sciences Group at Los Alamos National Laboratory (email: *gdwilson@lanl.gov*). LA-UR-04-6564.

1 Introduction

At 8:40p.m. on February 25, 1991, parts of an Iraqi Scud missile destroyed the barracks housing members of the United States Army's 14th Quartermaster Detachment. This was the single, most devastating attack on U. S. forces during the First Gulf War: 29 soldiers died and 99 were wounded. In the aftermath of this attack, there has been great focus on developing air defense systems capable of defending against ballistic missile attacks. The Critical Measurements and Counter Measures Program (CMCM), run by the United States Army Space and Missile Defense Command, conducts exercises to replicate projected ballistic missile threats. These exercises help the U. S. military collect realistic data to evaluate potential defensive measures. The high-fidelity hardware and realistic scenarios created for the exercises provide extensive optical, radar, and telemetry data (U. S. Army SMDC 2004).

CMCM is organized into campaigns. Each campaign chooses a new ballistic missile threat and develops two to four high fidelity launch vehicles that emulate the threat as closely as possible given intelligence information. While there is some reuse across campaigns, each set of launch vehicles is essentially a complex, one-of-a-kind, one-time-use system built for a specific data collection purpose. Typically, due to cost and schedule constraints, there are no "risk reduction" flights performed, so there is no full-system checkout before the actual flights. The systems are designed and built in a distributed fashion, with scientists and engineers from different companies designing, building, and integrating various parts of the vehicle. These campaigns are expensive (millions of dollars) and politically high profile.

The issue that we address in this paper is how to determine a preflight probability of mission success and how to assess areas of risk to the flight. Since there are no full-system tests, this involves careful system modeling and the integration of as much component, historical, and engineering data as possible. The applied problem described here is large and complex. The system itself is well-understood in some dimensions by the groups working on the project, but not in terms of its overall reliability and performance. Knowledge of the total system is distributed across two primary research and development contractors and several subcontractors, all of which are located in different parts of the country. Each research group understands its area of responsbility at a granular level, and there is working knowledge of how to build a missile that will fly, but the project teams do not have methodology or tools to assess or predict full-system performance or reliability.

There is heterogeneous data that explains different aspects of component and subcomponent performance, but very little sense of how that data relates or how to sensibly combine the data and propagate reliability estimates and their uncertainties to understand overall system reliability. There are hundreds of components and subcomponents that all perform differently. This is not a consulting problem where the client comes in with a single set of data, and the statistician immediately intuits the correct model and directly applies statistical tools. Our approach to grappling with this problem was to first build a qualitative model of the problem space (its parts and relationships) and then to migrate that qualitative model to a graphical statistical model. We used ethnographic interviewing and observation techniques (Meyer and Booker 2001) to elicit the problem structure, which was then represented in conceptual graphs (Sowa 2000). The ethnographic methods used to represent the model structure are described in detail in a separate article (McNamara and Leishman 2004). The framework that we use to quantitatively model the system and integrate the data is Bayesian networks (BN).

There is limited literature on the use of BNs in failure modes and effects analysis (Lee 2001) and reliability (summarized in Sigurdsson et al. (2001)), although there is quite a broad literature on using BNs for probabilistic modeling (e.g., Spiegelhalter (1998); Neil et al. (2000); Laskey and Mahoney (2000); Jensen (2001)). The reliability literature focuses primarily on discrete random variables; we extend this to the case where there are uncertainties in the conditional probability table. We apply the ideas from the general BN literature to the development of the system model and inference.

This paper will describe our approach to the assessment of Campaign 4 of the CMCM program (CMP-4). Section 2 details the development of the system representation using both qualitative and quantitative methods. Section 3 discusses the statistical model for the system and the information and data available to populate the model. Section 4 shows how the information was combined to make estimates. Section 5 contains conclusions and discussion.

2 Representing the System

2.1 Defining Mission Success

The questions of interest for CMP-4 included assessing the probability of mission success and identifying areas of technical risk. For large collaborative technical efforts to develop specialized technology, it is often difficult for those involved to have an integrated view of the project space, its goals, and the metrics for success. Likewise, at the beginning of the CMP-4 project, there was no clear definition of overall mission success. (The joke was "A mission is successful if I can write a press release that saves my job.") There were clearly some things that would make a mission unsuccessful—for example, if the vehicle explodes on the launch pad. However, there was no clear enumeration of intermediate negative events that would render a mission more or less successful, nor a clear understanding of how certain events in combination would lead to undesirable outcomes.

To develop a definition of mission success, we worked with experts at CMCM to understand what events occur to make up the mission. Over several months we also worked with the experts at the two main contractors on the project to understand what functions the missile had to perform for the mission events to transpire successfully. Furthermore, we worked with them to identify the parts of the system that had to correctly perform those functions. This events-functions-parts approach to modeling the technical system allowed us to characterize both the overall probability of mission success and to localize the areas of risk across the system. Unacceptably high risk or uncertainty could be traced to the event, function, part, or interaction of parts that was the root cause.

For CMP-4, the events that make up the mission fall into three categories: the threat-representative flight, the data collection, and the auxiliary experiments. Failure in any of these categories cause a mission to be unsuccessful. Figure 1 summarizes the events that make up the mission. The threat-representative events are on the left side of the diagram. Notice that there are nine different data collection streams; some start immediately after ignition, and others start after later events.

Once the mission was defined in terms of the events that make it up, then the question of mission success could be revisited. One issue that had to be addressed was whether mission success was a discrete or continuous quantity. It was decided that mission success could be defined as catastrophic failure (RED), degraded (YELLOW), or nominal (GREEN). This language was natural for the CMCM staff and contractors working on the program, as it is commonly used in the Department of Defense to describe categories of outcomes in technical and military missions. Each of the events in Figure 1 was also defined to have RED, YELLOW, or GREEN states. Table 1 summarizes what event states can cause catastrophic mission failure (RED); Table 2 summarizes the event states that cause a degraded mission (YELLOW).

For the events in Tables 1 and 2, each is contingent upon the previous events (from Figure 1) not failing catastrophically. For example, ignition fails catastrophically when the vehicle blows up on the launch pad. If the vehicle blows up before launch, all other events do not occur. The conditional specification of relationships in the mission suggests that a BN might be the appropriate representation for the probability of mission success.

2.2 Constructing Bayesian Networks

Formally, a BN is a pair $N = \langle (V, E), P \rangle$, where (V, E) are the nodes and edges of a directed acyclic graph and P is a probability distribution on V. Each node contains a random variable, and the directed edges between them define conditional dependences/independences among the random variables. Figure 2 summarizes the three probabilistic relationships that can be specified in a BN. More informally, a BN is useful when the structure of the model and information is "local," meaning it can be specified as depending on a "few" other variables and affecting a "few" others.

The conditional dependence/independence structures in Figure 2 may be useful in developing the network, for example, when the network represents a hierarchical Bayesian model. There are other heuristics that can be used to construct a BN. Neil et al. (2000) identifies five *idioms* or patterns that appear frequently in BNs. Of these, three appear in the CMP-4 system representation. The first is the definitional/synthesis idiom. This idiom captures the idea of "saying what something is": the child node is defined by its parents. For exam-

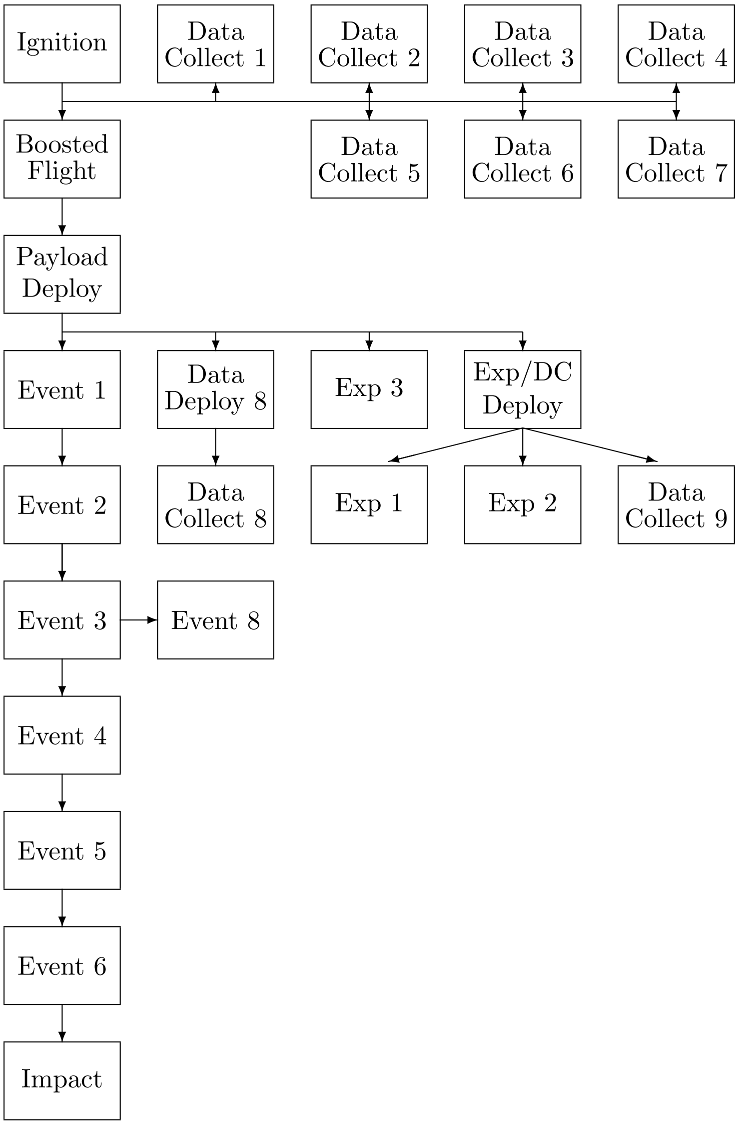


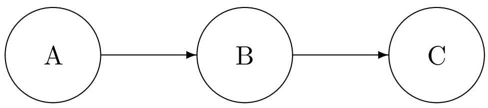
Figure 1: Mission Events

Table 1: Mission Success RED

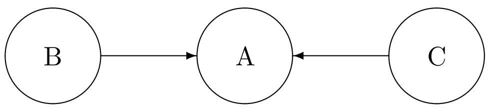
Event	State
Ignition	RED
Boosted Flight	RED
Payload Deploy	RED
Event 3	RED

Table 2: Mission Success YELLOW

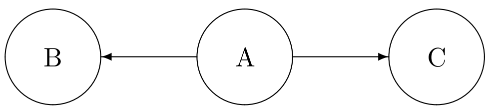
Event	State
Boosted Flight	YELLOW
Data Collection 1	RED
Data Collection 4	RED
Data Collection 5	RED
Data Collection 7	RED
Data Collection 8	RED
Data Collection 8	YELLOW
Data Collection 9	RED
Experiment 1	RED
Event 1	RED
Event 2	RED
Event 4	RED
Event 5	RED



(a) Serial: $\mathbf{P}(A, B, C) = \mathbf{P}(C|B)\mathbf{P}(B|A)\mathbf{P}(A)$



(b) Converging: $\mathbf{P}(A, B, C) = \mathbf{P}(A|B, C)\mathbf{P}(B)\mathbf{P}(C)$



(c) Diverging: $\mathbf{P}(A, B, C) = \mathbf{P}(C|A)\mathbf{P}(B|A)\mathbf{P}(A)$

Figure 2: Structures in a Bayesian Network

ple, in Figure 3a, velocity = distance/time. Since velocity is not random given distance and time, it technically does not have to be represented in the BN, but including it can clarify the model.

The second useful idiom is cause/consequence. This idiom captures the idea of "causal reasoning based on production or transformation": the parent nodes are the inputs to a process, and the child node is the output. For example, in Figure 3b, wire breaking or battery failing causes the power to fail. The cause/consequence idiom is ordered chronologically, with the parent nodes (inputs) occurring before the child node (outputs).

The third useful idiom is induction. This idiom captures the idea of "statistical and analogical reasoning using historical cases to say something about an unknown cause." This idiom is the basic model for Bayesian inference. In Figure 3c, a historical attribute (perhaps a population parameter) and a measure of similarity to the historical data are used to make a forecast.

In addition to the Neil et al. (2000) idioms, another probability structure that can be captured by a BN is a fault tree (Bobbio et al. 2001). The fault tree translation to a BN is like definitional/synthesis idiom, with two basic events that contribute to an intermediate event represented as two parents and a child. Figure 4 shows the correspondence between a fault tree AND gate and a BN AND node. Notice that a fault tree implies specific conditional probabilities.

3 System Model

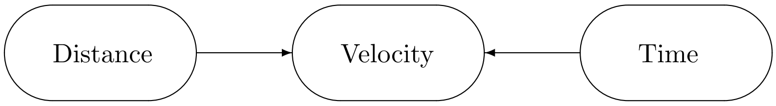
The joint distribution of V, the set of nodes in a BN, is given by

$$\prod_{v \in V} \mathbf{P}(v|\text{parents}[v]), \tag{1}$$

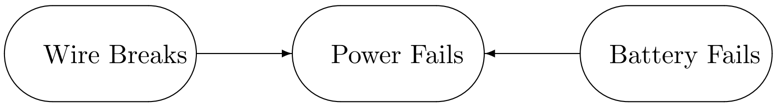
where the parents of a node are the set of nodes with an edge pointing to the node. For example, in the serial structure in Figure 2a, the parent of node C is node B, and node A has no parents.

Equation 1 shows that the joint distribution of the nodes in the BN is determined by a set of conditional distributions. For example, in Figure 1, one of the probabilities that needs to be assessed to determine the joint distribution of all of the events is $\mathbf{P}(\text{Data Collect 1} = \text{RED}|\text{Ignition} = \text{GREEN})$. Notice that the conditional dependence/independence structure of the BN greatly decreases the total number of probabilities that have to be specified. If the random variables are discrete and there is no conditional structure, then every possible combination of values of the random variables must be assessed.

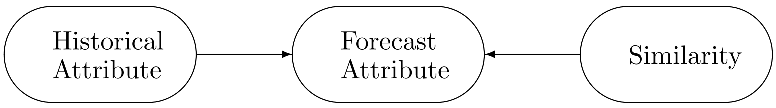
Consider again Figure 1 and Tables 1 and 2. These summarize the events that make up the mission and the states of these events that define mission success. To make a quantitative assessment of the probability of mission success, all of the conditional probabilities in Figure 1 need to be assessed. For CMP-4, these probabilities could not be elicited directly, nor were test data collected that addressed the probabilities directly. Consequently, once mission success



(a) Definitional/Synthesis Idiom

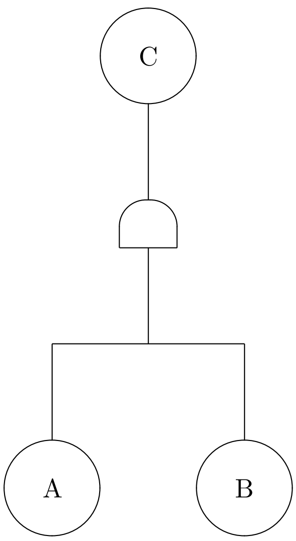


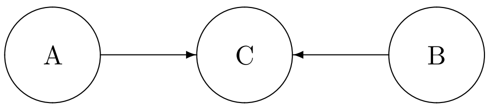
(b) Cause/Consequence Idiom



(c) Induction Idiom

Figure 3: Bayesian Network Idioms





$$P(C = 1|A, B) = I_{AB}$$

Figure 4: Fault Tree and Bayesian Network

was defined, the definition process began again for each of its component probabilities. This is where the events-functions-parts approach to modeling the system is relevant.

Consider, for example, the event boosted flight. Boosted flight can be decomposed into the BN given in Figure 5. In particular, boosted flight is made up of a series of functions. Functions correspond to the idea of network *fragments*, which are small groups of related variables that help structure the BN (Neil et al. 2000; Laskey and Mahoney 2000).

Again, these functions are not at the right granularity, as there is no data or information about the conditional probabilities. Figure 6 is the BN for roll control, which is a further decomposition of part of Figure 5. The boxed nodes are parts/physical components of the system; the circled nodes are functions or events. The conditional probabilities for the parts can be estimated from existing data for those parts, many of which were used in past missions. The conditional probabilities can also be elicited using standard expert judgment elicitation techniques, which is what we did for newer parts. Because parts make up functions, and functions comprise events, the conditional probabilities for the corresponding functions and events can then be calculated. This process was completed for the entire set of events in Figure 1 and resulted in a BN with approximately 600 nodes.

If all of the random variables V are discrete and all of the conditional probabilities in Equation 1 are specified by point values, then there are exact algorithms to compute the joint distribution for V and marginal distributions for any subset of variables (Spiegelhalter et al. 1996; Jensen 2001). However, for CMP-4, the conditional probabilities could not be elicited as point values.

The analysis of the CMP-4 was conducted along with the design, construction, and testing of the system itself. Consequently, at various points during the analysis, there was no "data" available, at least in terms of repeated observations of particular parts, functions, or events. To begin the analysis, Figure 1 was decomposed, following the example of Figures 5 and 6. Each conditional probability required to fully specify the joint distribution of the resulting Bayesian network was elicited from subject matter experts. Most of the experts had very local knowledge—they could provide information for a few nodes conditional on their parents.

The experts were unable to provide precise conditional probabilities, however, and they did not have access to historical data that they could use to formally estimate the probabilities. Consequently, values were elicited within the ranges given in Table 3. The experts were asked to identify failure modes for each part, and were then asked to estimate the chance of part failure. We reviewed the various biases associated with estimating probabilities with the respondents and asked them to remain aware of bias issues throughout the discussions; in addition, they were asked to justify their estimates with reference to other experiments or tests. Many of the interviews were conducted in groups, which also helped to alleviate bias.

As examples, some of the elicited information for Figure 6 had the following form:

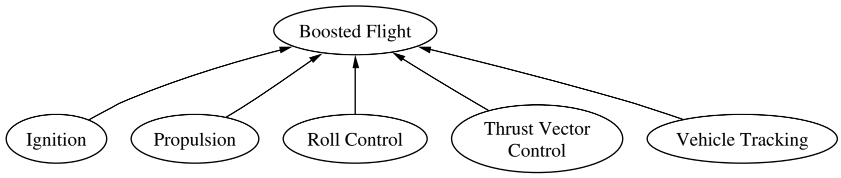


Figure 5: Boosted Flight Bayesian Network

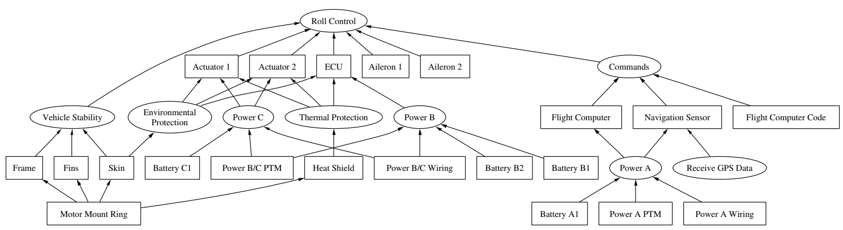


Figure 6: Roll Control Bayesian Network

- The flight computer code is RED with $\Phi = 2$
- If the navigation sensor fails RED, commands are YELLOW with $\Phi = 3$
- If the flight code is RED, then commands are RED
- If Power B is RED, the electronics control unit is RED
- If environmental protection is RED, the electronics control unit is RED with $\Phi = 4$.

4 Information Combination

Statistically, we have the following problem. We want to understand $p_S = \mathbf{P}(\text{Mission Success})$. We have very little information about p_S directly, and through an intensive modeling process, we have been able to rewrite $p_S = f(p_1, \ldots, p_n)$, where the $f()$ and the p_i are defined by a Bayesian network structure. For all of the p_i, we have elicited distributions. In addition, for some of the p_i (and functions of the p_is) we have additional data. We would like to calculate $\pi(p_1, \ldots, p_n|\text{Data})$ or at least draw a random sample from it. We can then use Monte Carlo to calculate $p_S = f(p_1, \ldots, p_n)$ for each random draw from $\pi()$ and thus calculate the distribution of p_S. Even if we cannot write down $f()$ explicitly, there are exact algorithms to compute it (Spiegelhalter et al. 1996; Jensen 2001).

Specifically, in addition to the elicited probabilities, there was some historical data available for the entire mission: in seven previous flights, two were RED, three were YELLOW, and two were GREEN. There was also some data available on the lifetimes of electronic components under nominal conditions. The engineers viewed their system tests and simulations as providing state information about certain events or functions. The individual parts within the tail section subsystem performed well during testing, and therefore the engineering assessment for the function Vehicle Stability was GREEN.

Calculating $\pi(p_1, \ldots, p_n|\text{Data})$ is straightforward if the p_i are independent and the data relates to only one p_i. If data is available about functions of the p_i, for example about historical mission success, then the method proposed in Hamada et al. (2004) and Johnson et al. (2003) can be used: substitute $f(p_1, \ldots, p_n)$ for p_S in the likelihood.

As an example, the Bayesian network in Figure 6 was evaluated. For elicited interval values, the left and right endpoints were used, with 0.001 as the left endpoint for $\Phi = 1$ and 0.999 as the right endpoint for $\Phi = 5$. The marginal probabilities for roll control using the left endpoints are

$$\begin{aligned}
\mathbf{P}(\text{Roll Control} = \text{RED}) &= 0.038 \\
\mathbf{P}(\text{Roll Control} = \text{YELLOW}) &= 0.039 \\
\mathbf{P}(\text{Roll Control} = \text{GREEN}) &= 0.922.
\end{aligned}$$

The marginal probabilities for roll control using the right endpoints are

$$\begin{aligned}
\mathbf{P}(\text{Roll Control} = \text{RED}) &= 0.355 \\
\mathbf{P}(\text{Roll Control} = \text{YELLOW}) &= 0.216 \\
\mathbf{P}(\text{Roll Control} = \text{GREEN}) &= 0.429.
\end{aligned}$$

If testing showed that vehicle stability could be assumed to be GREEN, then the "left" probabilities change to $(0.013, 0.009, 0.977)$ and the "right" probabilities change to $(0.201, 0.129, 0.670)$.

Table 3 contains beta distributions that could be used to capture the elicited intervals. These were chosen so that the endpoints of the intervals are approximately the 5th and 95th quantiles of the beta distribution. These parameters of the distributions can be calculated using the formulas in Mosleh and Apostolakis (1982) or by grid search in R (The R Project for Statistical Computing 2004). Figure 7 shows the marginal posterior distributions for the probability that roll control is RED using the uniform and beta distributions. Table 4 summarizes the 95% credible intervals for mission success for the uniform and beta distributions; the third and fourth entries are the credible intervals when vehicle stability is assumed to be GREEN. The probabilities are shifted left slightly when the beta distributions are used, but the qualitative conclusions remain unchanged.

5 Conclusions

The final Bayesian network for CMP-4 contained approximately 600 nodes. Neil et al. (2000) summarizes many of the issues that surround working with a model of this size:

> Large knowledge-based systems, including BNs, are subject to the same forces as any other substantial engineering undertaking. The customer might not know what they want; the knowledge engineer may have difficulty understanding the domain; the tools and methods applied may be imperfect; dealing with multiple ever-changing design abstraction is difficult, etc. In the end these issues, along with people, economic and organizational factors, will impact on the budget, schedule, and quality of the end product. (p. 265)

There are inherent challenges in working on a project like this. For the people in charge of such a system, even articulating an overarching definition for success can be difficult—getting a workable statistical model out of it, even more so. The authors of this paper represent a collaboration between statisticians and ethnographers with expertise in eliciting model structures and expert judgment. They worked together to develop first a qualitative model of the events, functions, and parts of the CMP-4 system, and second a graphical statistical model that added elicited and available data to provide quantitative answers for CMP-4 managers.

Table 3: Elicited Probability Ranges

Elicited Value (Φ)	Range	Distribution
0	0	
1	$(0, 0.01]$	Beta(2.5,550)
2	$[0.01, 0.1]$	Beta(2.5,55)
3	$[0.1, 0.25]$	Beta(10,50)
4	$[0.25, 0.5]$	Beta(15,25)
5	$[0.5, 1)$	Beta(5,1.25)
6	1	

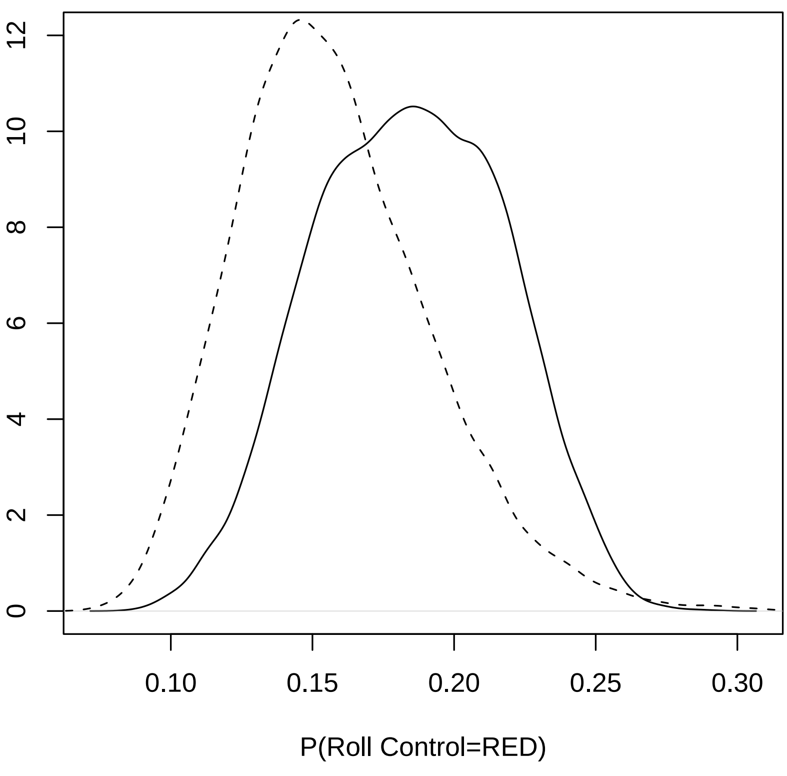


Figure 7: Marginal Probabilities for Roll Control

There are no end-to-end computer tools that help with system representation, statistical model formulation, and inference. However, at Los Alamos National Laboratory (LANL), we are working on tools for system representation (Gromit; see Koehler and Klamann 2004) and inference (YADAS; see Graves 2003). We hope that eventually these tools will be part of an end-to-end suite of tools.

This problem and these methods are of particular interest to the Department of Energy national laboratories. Since the U. S. ended its full-scale nuclear test program in 1992, LANL has been developing statistical methodology to certify the reliability and performance of the U. S. nuclear stockpile. The assessment issues are quite similar to those faced by the CMCM program: no full-system testing, and the need to develop complete system models and integrate all available information and data. We believe that these methods are applicable to a variety of complex systems and that they can provide a traceable and defensible estimate of system metrics, which can facilitate other planning and problem-solving efforts.

References

[1] A. Bobbio, L. Portinale, M. Minichino, E. Ciancamerla (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety* 71: 249-260.

[2] T. Graves (2003). YADAS: Yet Another Data Analysis System. http://yadas.lanl.gov.

[3] M. Hamada, H. F. Martz, C. S. Reese, T. Graves, V. Johnson, A. G. Wilson (2004). A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. *Reliability Engineering and System Safety* 86 (3), December 2004, pp. 297–305.

[4] F. V. Jensen (2001). Bayesian Networks and Decision Graphs. Springer-Verlag New York, New York.

[5] V. E. Johnson, T. L. Graves, M. S. Hamada, C. S. Reese (2003). A hierarchical model for estimating the reliability of complex systems. In *Bayesian Statistics 7*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, M. West. Oxford University Press, Oxford, UK.

[6] A. Koehler, R. Klamann (2004). Gromit: A tool for systems ethnography. LA-UR-04-4370, Los Alamos National Laboratory, Los Alamos, NM.

[7] K. B. Laskey, S. Mahoney (2000). Network engineering for agile belief network models. *IEEE Transactions on Knowledge and Data Engineering* 12(4): 487-498.

[8] B. H. Lee. Using Bayesian belief networks in industrial FMEA modeling and analysis. *Proceedings of the Annual Reliability and Maintainability Symposium.* International Symposium on Product Quality and Integrity, Philadelphia, PA, January 22-25, 2001, pp. 7-15.

[9] L. A. McNamara, D. A. Leishman (2004). Interlopers, translators, scribes, and seers: anthropology, knowledge representation and Bayesian statistics for predictive modeling in multidisciplinary science and engineering projects. In Multidisciplinary Studies of Visual Representations and Interpretations, ed. G. Malcolm. Elsevier Science, to appear.

[10] MB01 M. Meyer, J. Booker (2001). Eliciting and Analyzing Expert Judgment: A Practical Guide. ASA-SIAM, Philadelphia.

[11] A. Mosleh, G. Apostolakis (1982). Some properties of distributions useful in the study of rare events. *IEEE Transactions on Reliability* R-31(1): 87-94.

[12] M. Neil, N. Fenton, L. Nielsen (2000). Building large-scale Bayesian networks. *Knowledge Engineering Review* 15(3): 257-284.

[13] R04 The R Project for Statistical Computing (2004). http: www.r-project.org.

[14] J. H. Sigurdsson, L. A. Walls, J. L. Quigley (2001). Bayesian belief nets for managing expert judgement and modelling reliability. *Quality and Reliability Engineering International* 17: 181-190.

[15] J. Sowa (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing, Pacific Grove, CA.

[16] D. J. Spiegelhalter (1998). Bayesian graphical modeling: A case-study in monitoring health outcomes. *Applied Statistics* 47(1): 115-133.

[17] D. Spiegelhalter, A. Thomas, N. G. Best (1996). *Computation on Bayesian Graphical Models.* In Bayesian Statistics 5, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith. Oxford University Press, Oxford, UK, pp. 407-426.

[18] United States Army Space and Missile Defense Command Technical Center (2004). Critical Measurements & Countermeasures Program Fact Sheet. http://www.smdc.army.mil/FactSheets/CMP.pdf.

Table 4: Mission Success 95% Credible Intervals

Distribution	Mission Success RED	Mission Success YELLOW
Uniform	$(0.12, 0.25)$	$(0.083, 0.18)$
Beta	$(0.10, 0.24)$	$(0.074, 0.18)$
Uniform with VS = G	$(0.059, 0.13)$	$(0.037, 0.091)$
Beta with VS = G	$(0.048, 0.11,)$	$(0.031, 0.087)$